

HP-10010078

Patent

UNITED STATES PATENT APPLICATION

for

A METHOD FOR VISUALIZING LARGE VOLUMES OF MULTIPLE-
ATTRIBUTE DATA WITHOUT AGGREGATION USING A PIXEL BAR CHART

Inventors:

Ming C. Hao
Umeshwar Dayal
Meichum Hsu
Daniel Kiem
Julian Ladisch

prepared by:

WAGNER, MURABITO & HAO LLP

Two North Market Street

Third Floor

San Jose, CA 95113

(408) 938-9060

09047393 072011

A METHOD FOR VISUALIZING LARGE VOLUMES OF MULTIPLE-
ATTRIBUTE DATA WITHOUT AGGREGATION USING A PIXEL BAR CHART

FIELD OF INVENTION

5 The present invention relates to the field of data visualization.
Specifically, the present invention relates to a method for visualizing large
volumes of data having multiple attributes without aggregation.

BACKGROUND OF THE INVENTION

10

Modern organizations and corporations often rely on computer
systems to manage the massive amounts of data acquired in the course of
running their operations. Interpreting large volumes of data often presents
great challenges to those responsible for understanding the data. A
15 common method for visualizing large volumes of data is to use a bar chart.
A bar chart is a type of graph in which different values are represented by
rectangular bars.

In a typical bar chart, each bar represents a single data value. It is
20 common to represent aggregated data using two-dimensional or three-
dimensional bar charts. When using a bar chart, the number of data items
shown at the same time must be pre-decided. Additionally, the bar chart
can only show a small range of data items (10-20), because showing too

many items will obscure the data. Furthermore, if a three-dimensional bar chart is used, some bars hide others and overlapping makes the visual impression misleading, as the area seen is not proportional to the value to be represented. In these cases, valuable information often gets lost.

5

Figure 1 illustrates a three-dimensional bar chart 100 of over 600,000 customer orders and corresponding purchase analysis. To allow for viewing, all data is aggregated by order 110, price 120 and quantity 130. No individual customer information is shown, only the aggregate total of customer information. The display is difficult to understand, due to some data (e.g., column 140) being disproportionately larger than other columns, making it difficult to understand the values of the smaller columns. Furthermore, some columns (e.g., column 140) hide other columns, making that data extremely difficult, if at all possible, to understand.

15

Another common method used to visualize large volumes of data is a stack chart 200 as shown in Figure 2. A stack chart is similar to a bar chart, but also allows for the area of a color to be proportional to the value of a particular set of data.

20

Stack chart 200 illustrates an exemplary stack chart illustrating sources of energy in the United States by year. The area of each color is proportional to the amount of energy provided by the respective source in

that year. Each year is a group. Only five groups are shown in Figure 2.

Limiting the number of groups to five allows for easy viewing of the data for the selected years, but severely limits the amount of data displayed. It may be desirable to view more data than just the five years displayed.

- 5 Increasing the number of groups, for example by showing all 372 months from 1950 until 1980, results in an overwhelming amount of data that is difficult to understand and interpret. If stack chart 200 were displayed on a standard computer system display of 1024x768 pixels, the groups would have a width of three pixels. These narrow groups do not allow the showing
- 10 of small width variations accurately like those from 1970 until 1980 (e.g., 2.4, 3.1, 2.7) in stack chart 200. In addition, at least 25% of the screen space is wasted for gaps between the groups, even if only one pixel is used as a gap.

- 15 Due to constraints dictated by the viewing limits of human eye, both bar charts and stack charts are required to reduce the number of values shown as the number of records increases. Otherwise, the chart obscures the nature of the data, as the chart will contain too much data in too small a space for a human reader to process and understand. The number of
- 20 values can be reduced by limiting the number of groups showing or by aggregating data. Both selection and aggregation reduce the information seen, limiting access to more detailed data.

[illegible]

SUMMARY OF THE INVENTION

The present invention provides a method for visualizing large volumes of data having multiple attributes without requiring aggregation of the data. A method for graphically presenting and visually mining large volumes of data using a graphically displayable array is presented. In one embodiment, the graphically displayable array is a pixel array. Data comprising a plurality of records is received, wherein each record has multiple attributes. A first attribute, second attribute and third attribute are selected from the plurality of records. In one embodiment, the visual indicator is a color selected from a range of colors. In one embodiment, the third attribute is the same attribute as the first attribute. In another embodiment, the third attribute is the same attribute as the second attribute. In one embodiment, the attributes selected to construct a graphically displayable array are predetermined. In another embodiment, a user selects the attributes. The plurality of records are arranged to construct a graphically displayable array, wherein the graphically displayable array comprises a plurality of pixels or data points. Each of pixel or data point represents one record of the plurality of records wherein the first attribute corresponds to a first axis, the second attribute corresponds to a second axis, and the third attribute corresponds to a visual indicator.

BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying drawings, which are incorporated in and form a part of this specification, illustrate embodiments of the invention and,
5 together with the description, serve to explain the principles of the invention:

FIGURE 1 illustrates a three-dimensional bar chart in accordance with the prior art.

10 FIGURE 2 illustrates a stack chart in accordance with the prior art.

FIGURE 3 illustrates an exemplary computer system on which embodiments of the present invention may be practiced.

15 FIGURE 4 is a flowchart diagram illustrating steps in a process for graphically presenting large volumes of data in accordance with an embodiment of the present invention.

20 FIGURE 5a is a block diagram illustrating the sorting of a plurality of records by a first attribute and dividing the records into groups accordingly in accordance with an embodiment of the present invention.

FIGURE 5b is a block diagram illustrating the sorting of a plurality of records within a group by a second attribute in accordance with an embodiment of the present invention.

- 5 FIGURE 5c is a block diagram illustrating the sorting of a plurality of records within a horizontal line of a group by a third attribute in accordance with an embodiment of the present invention.

- 10 FIGURE 6a illustrates a graphically displayable array wherein the third attribute (associated with color) is the same as the first attribute (associated with the horizontal axis) in accordance with an embodiment of the present invention.

- 15 FIGURE 6b illustrates a graphically displayable array wherein the third attribute (associated with color) is the same as the second attribute (associated with the vertical axis) in accordance with an embodiment of the present invention.

- 20 FIGURE 6c illustrates a graphically displayable array wherein the third attribute (associated with color) is different than both the first attribute (associated with the horizontal axis) and the second attribute (associated with the vertical axis) in accordance with an embodiment of the present invention.

DETAILED DESCRIPTION

The application file contains at least one drawing executed in color. Copies of this patent application publication with color drawings will be provided by the Office upon request and payment of the necessary fee.

In the following detailed description, for purposes of explanation, numerous specific details are set forth in order to provide a thorough understanding of the present invention. However, it will be apparent to one skilled in the art that the present invention may be practiced without these specific details. In other instances, well-known structures and devices are not described in detail in order to avoid obscuring aspects of the present invention.

Some portions of the detailed descriptions which follow are presented in terms of procedures, steps, logic blocks, processing, and other symbolic representations of operations on data bits within a computer memory. These descriptions and representations are the means used by those skilled in the data processing arts to most effectively convey the substance of their work to others skilled in the art. A procedure, computer executed step, logic block, process, etc., is here and generally conceived to be a self-consistent sequence of steps of instructions leading to a desired result. The steps are those requiring physical manipulations of data

representing physical quantities to achieve tangible and useful results. It has proven convenient at times, principally for reasons of common usage, to refer to these signals as bits, values, elements, symbols, characters, terms, numbers or the like.

5

It should be borne in mind, however, that all of these and similar terms are to be associated with the appropriate physical quantities and are merely convenient labels applied to these quantities. Unless specifically stated otherwise as apparent from the following discussions, it is

10

appreciated that throughout the present invention, discussions utilizing terms such as "receiving", "sorting", "constructing", "interacting", "placing" or the like, refer to the actions and processes of a computer system, or similar electronic computing device. The computer system or similar electronic device manipulates and transforms data represented as electronic

15

quantities within the computer system's registers and memories into other data similarly represented as physical quantities within the computer system memories into other data similarly represented as physical quantities within the computer system memories or registers or other such information storage, transmission, or display devices.

20

With reference to Figure 3, portions of the present invention are comprised of computer-readable and computer executable instructions which reside, for example, in computer-usable media of a computer

system. Figure 3 illustrates an exemplary computer system 300 on which
embodiments (e.g., process 500 of Figure 5 and process 600 of Figure 6) of
the present invention may be practiced. It is appreciated that computer
system 300 of Figure 3 is exemplary only and that the present invention can
5 operate within a number of different computer systems including general
purpose computer systems, embedded computer systems, and stand
alone computer systems specially adapted for controlling automatic test
equipment.

10 Computer system 300 includes an address/data bus 310 for
communicating information, a central processor 301 coupled with bus 310
for processing information and instructions, a volatile memory 302 (e.g.,
random access memory RAM) coupled with the bus 310 for storing
information and instructions for the central processor 301 and a non-volatile
15 memory 303 (e.g., read only memory ROM) coupled with the bus 310 for
storing static information and instructions for the processor 301.

Computer system 300 also includes a data storage device 304 ("disk
subsystem") such as a magnetic or optical disk and disk drive coupled with
20 the bus 310 for storing information and instructions. Data storage device
304 can include one or more removable magnetic or optical storage media
(e.g., diskettes, tapes) which are computer readable memories. Memory
units of system 300 include volatile memory 302, non-volatile memory 303

and data storage device 304. In one embodiment, volatile memory 302 is partitioned to comprise a number of distinct, independently operating memory units.

5 Computer system 300 can further include an optional signal generating device 308 (e.g., a modem, or a network interface card "NIC") coupled to the bus 300 for interfacing with other computer systems. Also included in computer system 300 of Figure 3 is an optional alphanumeric input device 106 including alphanumeric and function keys coupled to the
10 bus 310 for communicating information and command selections to the central processor 301. Computer system 300 also includes an optional cursor control or directing device 307 coupled to the bus 310 for communicating user input information and command selections to the central processor 301. An optional display device 305 can also be coupled
15 to the bus 310 for displaying information to the computer user. Display device 305 may be a liquid crystal device, other flat panel display, cathode ray tube, or other display device suitable for creating graphic images and alphanumeric characters recognizable to the user. Cursor control device 307 allows the computer user to dynamically signal the two dimensional
20 movement of a visible symbol (cursor) on a display screen of display device 305. Many implementations of cursor control device 307 are known in the art including a trackball, mouse, touch pad, joystick or special keys on alphanumeric input device 306 capable of signaling movement of a given

direction or manner of displacement. Alternatively, it will be appreciated that a cursor can be directed and/or activated via input from alphanumeric input device 306 using special keys and key sequence commands.

5 Figure 4 is a flowchart diagram illustrating steps in a process 400 for graphically presenting and visually mining large volumes of data having multiple attributes without requiring aggregation of the data, in accordance with an embodiment of the present invention. Steps of process 400, in the present embodiment, may be implemented with any computer languages
10 used by those of ordinary skill in the art. In one embodiment, process 400 is for graphically presenting and visually mining large volumes of data using a graphically displayable array. In one embodiment, the graphically displayable array is a pixel array.

15 At step 410 of process 400, data comprising a plurality of records is received. Each record comprises a plurality of attributes, wherein each attribute corresponds to a particular piece of information of a record. For example, a consumer electronics business has a record for each order the business handles. In this example, each record may have attributes
20 corresponding to the order number, the price of the order, the customer identification number (ID), and the quantity of items ordered.

At step 420, it is determined which attributes will be selected for inclusion in the construction of a graphically displayable array. In one embodiment, a first attribute of the plurality of attributes is selected, wherein the first attribute corresponds to a horizontal axis of an array. A second
5 attribute of the plurality of attributes is selected, wherein the second attribute corresponds to a vertical axis of the array. A third attribute of the plurality of attributes is selected, wherein the third attribute corresponds to a color.

It should be appreciated that the third attribute is selected from the
10 entire plurality of attributes. In one embodiment, the third attribute is the same attribute as that selected as the first attribute. In another embodiment, the third attribute is the same attribute as that selected as the second attribute.

15 At step 430, the plurality of records are sorted by the first attribute. The records are then divided into groups by the first attribute, such that records that have the same value for the first attribute constitute a group.

In one embodiment, each record is represented by one unique pixel
20 of a display. Each group has to have at least as many records as the height of the array in pixels. The height of the array is predetermined, based on user inputs. In one embodiment, the height of the array is determined as a function of the number of vertical pixels comprising the display and the total

number of records comprising the array. The number of records in each group determines the width of each group of the array. The area of each group is proportional to the number of records in each group.

5 In another embodiment, each record is represented by a data point of a display, wherein the data point comprises a plurality of pixels. Each group has to have at least as many records as the height of the array in pixels or data points.

10 Figure 5a is a block diagram 500 illustrating the sorting of a plurality of records by a first attribute and dividing the records into groups accordingly in accordance with an embodiment of the present invention. Diagram 500 illustrates an exemplary array comprising a horizontal axis 502 and a vertical axis 504. Each record is sorted according to the first attribute and
15 divided into groups. Group one 506 consists of records all having the same value for the first attribute. Similarly, group two 508, group three 510 and group four 512 consist of records with identical values of the first attribute, respectively. The arrows indicate the width of each group, respectively.

20 At step 440, the records are sorted by the second attribute within each group. In one embodiment, the records are sorted from the lowest value of the second attribute to the highest value of the second attribute. The record having the lowest value of the second attribute is placed in the

bottom left of each group, and records are placed from left to right, moving up the group, until all records have been placed within the group.

Figure 5b is a block diagram 520 illustrating the sorting of a plurality
5 of records within a group by a second attribute in accordance with an
embodiment of the present invention. Diagram 530 illustrates an exemplary
array comprising a horizontal axis 502 and a vertical axis 504. Each record
is sorted according to the second attribute and within each group. The
arrows indicate the possible orders of sorting the records by the second
10 attribute. In one embodiment, the records are sorted vertically, and ordered
in horizontal lines from left to right.

At step 450, the records are sorted by the third attribute within each
horizontal line of each group. In one embodiment, each horizontal line is
15 sorted such that the lowest value to the highest value is sorted from left to
right within the horizontal line.

Figure 5c is a block diagram 540 illustrating the sorting of a plurality
of records within a horizontal line of a group by a third attribute in
20 accordance with an embodiment of the present invention. Diagram 540
illustrates an exemplary array comprising a horizontal axis 502 and a
vertical axis 504. Each record is sorted according to the third attribute and
within each horizontal line of each group. In one embodiment, the records

are ordered in the horizontal lines from left to right. The arrows indicate the possible orders of sorting the horizontal lines.

At step 460, each record is assigned a color corresponding to the value of the third attribute. It should be appreciated that any attribute may be selected as the third attribute for purposes of the present invention. In one embodiment, the color is calculated from the value of the third attribute. In one embodiment, non-linear 256 RGB (red-green-blue) color scale is used for determining the color for each record. In another embodiment, a non-linear gray-scale color scale is used to determine the color for each record. It should be appreciated that any color scale or range, both linear and non-linear, may be used in regard to the present invention.

In one embodiment, the value of the third attribute is normalized to the range 0 to 1. In one preferred embodiment, the normalization is nonlinear. Then the range 0 to 1 is mapped to a color range.

At step 470, a graphically displayable array comprising the previously sorted records (e.g. sorting of steps 430 – 450) is constructed in the memory (e.g., volatile memory 302 of Figure 3) of a computer system.

At step 480, the graphically displayable array constructed in step 470 is drawn on a display device (e.g., display device 305 of Figure 3) of a

computer system. In one embodiment, the graphically displayable array is a pixel array (e.g., a pixel bar chart). In the present embodiment, each record is graphically presented in the display device as one pixel. Each record of a group is represented as one pixel. The area of a group
5 represents the number of pixels comprising the group.

In one embodiment, a user may interact with the graphically displayable array by moving a cursor to a pixel or data point to access the information of the represented record. A “drill down” technique allows for
10 the viewing of all related information after selecting a single record. In another embodiment, a user may view all related information for a cluster of pixels or data points by selecting an area of the array with a cursor.

In one embodiment, the attributes used for grouping horizontally, sorting vertically and for coloring can be selected and changed interactively
15 to allow for faster access to more valuable information.

Figures 6a, 6b and 6c illustrate a series of exemplary graphically displayable pixel arrays (e.g., pixel bar charts) wherein the third attribute is varied across each array. Figure 6a illustrates a graphically displayable
20 pixel array 600 wherein the third attribute is the same as the first attribute in accordance with an embodiment of the present invention. Pixel array 600 comprises horizontal axis 602 and vertical axis 604.

As described above, a plurality of records are first sorted by the first attribute and divided into groups according to the first attribute. In the present embodiment, the third attribute (e.g., the attribute associated with a corresponding color) is the same attribute selected as the first attribute. Color is used to represent the third attribute of a record. The color is calculated from the value of each attribute, and the color is assigned to each record accordingly.

In Figure 6a, as the third attribute is the same as the first attribute, each group will be comprised of records with the same third attribute. As a result, each record within each group is assigned the same color. In the present embodiment, every record of group 606 is assigned the color red, every record of group 608 is assigned the color orange, every record of group 610 is assigned the color yellow.

Figure 6b illustrates a graphically displayable pixel array wherein the third attribute is the same as the second attribute in accordance with an embodiment of the present invention. Pixel array 620 comprises horizontal axis 622 and vertical axis 624.

As described above, after the plurality of records are sorted by the first attribute and divided into groups, the records within each group are sorted

vertically by the second attribute. In the present embodiment, the third attribute (e.g., the attribute associated with a corresponding color) is the same attribute selected as the second attribute.

5 In Figure 6b, as the third attribute is the same as the second attribute, the vertical sorting of the records of each group are represented in color by the third attribute. As the values for the third attribute increase, the color assigned to each record changes to reflect the difference. The records of group 626, in the present embodiment, gradually change in color from red
10 to purple, based on the corresponding value of the third attribute of each record. Similarly, the records of group 628 and 630 also gradually change in color from red to purple.

It should be appreciated that the embodiments discussed above in
15 pixel array 600 of Figure 6a and pixel array 620 of Figure 6b are special situations where the third attribute selected is the same attribute as either the first or second attribute. Figure 6c illustrates a graphically displayable pixel array 640 wherein the third attribute is different than both the first attribute and the second attribute in accordance with an embodiment of the
20 present invention. Pixel array 640 comprises horizontal axis 642 and vertical axis 644.

Each horizontal line of each group (e.g., groups 646, 648, and 650) are sorted by a third attribute. The records of each horizontal line of each group gradually change in color from red to purple as a result of the corresponding value of the third attribute of the record.

5

In one embodiment of the present invention, a user may interact with the data in a number of ways. In one embodiment, the attributes used for grouping horizontally, sorting vertically and for the coloring (e.g., the first, second and third attributes) can be selected and changed interactively to allow faster identification of valuable information. The user can interactively change any of the attributes of the present pixel bar chart to get a set of new pixel arrays.

In one embodiment, each record comprises more than three attributes. In constructing the series of multiple linked graphically displayable arrays, the first attribute and the second attribute remain the same across all arrays. However, the third attribute can be changed interactively to any of the remaining attributes to allow access to different information.

20

In one embodiment, a "drill down" technique allows the viewing of all related information after picking a single record. A user interacting with a

cursor may select a single record. By selecting a record, the user can view all attributes related to the record.

The non-aggregation information visualization technique of the present invention provides solutions to meet the need of automatic data preparation for the visual data mining of massive data volumes. The present invention retains the simplicity of solutions for viewing small data volumes. Furthermore, the present invention effectively uses screen space to represent each record without cluttering the display, allowing a user to easily discover patterns and correlations. The present invention provides a visual impression by representing the value of a record by a color and representing the number of records by the area of a group. With "drill down" capability, a user can navigate through each record to find detail information. Each record is represented by one pixel, allowing millions of records to be displayed at the same time. Each individual record can be accessed interactively, by allowing direct access to the detail data by picking at single pixels.

The present invention also provides the advantage of representing each record by one pixel allowing millions of records to be displayed at the same time without aggregation (e.g., losing information items).

[illegible]